

PROBE DESIGNING METHOD AND BIOCHIP

Publication number: JP2002330768

Publication date: 2002-11-19

Inventor: UENO SHINGO; NAKASHIGE AKIRA; NOZAKI YASUYUKI; MATSUMOTO TOSHIKO; TAMURA TAKURO

Applicant: HITACHI SOFTWARE ENG

Classification:

- international: G01N33/53; C12M1/00; C12N15/09; C12Q1/68; G01N37/00; G01N33/53; C12M1/00; C12N15/09; C12Q1/68; G01N37/00; (IPC1-7): C12N15/09; C12M1/00; C12Q1/68; G01N33/53; G01N37/00

- European:

Application number: JP20010142170 20010511

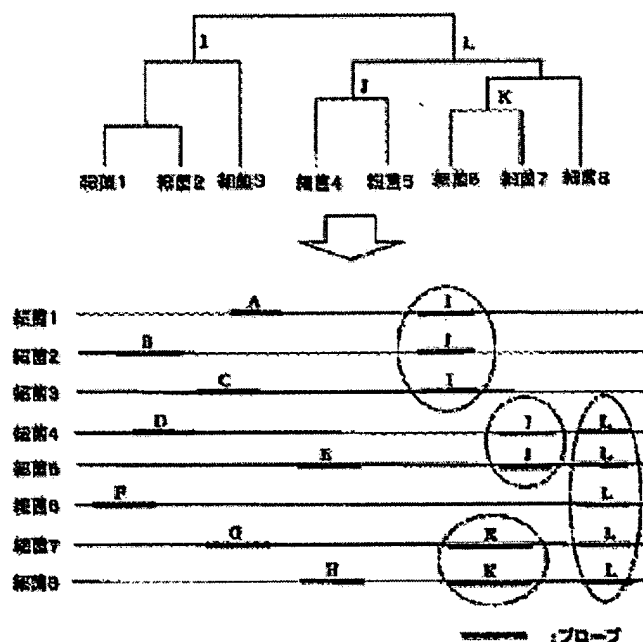
Priority number(s): JP20010142170 20010511

Report a data error here

Abstract of JP2002330768

PROBLEM TO BE SOLVED: To provide a method for rigorously discriminating the bio-species with a biochip and a method to enable the discrimination of bio-species at the level of species or above and to facilitate the selection of specific probe among various bio-species.

SOLUTION: When a base sequence common in leaves at a certain node or below is present on the genealogical tree of a target, a probe is designed by using the base sequence as the probe intrinsic to the node. The rigorous discrimination of bio-species and the discrimination of bio-species at the level of species or above can be carried out by using the probe and a probe intrinsic to the leaf as a set. Even if the base sequences corresponding to the leaf are the same or similar to each other, it can be used as a probe when the sequences corresponding to the node are different and, accordingly, the specific probe among various bio-species is easily selectable.



Data supplied from the esp@cenet database - Worldwide

Family list4 family members for: **JP2002330768**

Derived from 4 applications

[Back to JP2002330](#)**1 METHOD OF DESIGNING BIOCHIP AND PROBE****Inventor:** NOZAKI YASUYUKI; NAKASHIGE AKIRA; **Applicant:** HITACHI SOFTWARE ENG
(+3)**EC:** **IPC:** G01N33/53; C12M1/00; C12N15/09 (+18)**Publication info:** JP2002296279 A - 2002-10-09**2 PROBE DESIGNING METHOD AND BIOCHIP****Inventor:** UENO SHINGO; NAKASHIGE AKIRA; (+3) **Applicant:** HITACHI SOFTWARE ENG**EC:** **IPC:** G01N33/53; C12M1/00; C12N15/09 (+12)**Publication info:** JP2002330768 A - 2002-11-19**3 Biochip and method of designing probes****Inventor:** NOZAKI YASUYUKI (JP); UENO SHINGO **Applicant:**

(JP); (+3)

EC: C12Q1/68B10A **IPC:** C12Q1/68; C12Q1/68; (IPC1-7): C12Q1/68
(+2)**Publication info:** US2002160401 A1 - 2002-10-31**4 Biochip and method of designing probes****Inventor:** NOZAKI YASUYUKI (JP); UENO SHINGO **Applicant:** HITACHI SOFTWARE ENG (US)

(JP); (+3)

EC: C12Q1/68B10A **IPC:** C12Q1/68; C12Q1/68; (IPC1-7): C12P21/06
(+1)**Publication info:** US2004219593 A1 - 2004-11-04

Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-330768

(P2002-330768A)

(43) 公開日 平成14年11月19日(2002.11.19)

(51) Int.Cl. ⁷	識別記号	F I	サーチト*(参考)
C 1 2 N 15/09	Z N A	C 1 2 M 1/00	A 4 B 0 2 4
C 1 2 M 1/00		C 1 2 Q 1/68	A 4 B 0 2 9
C 1 2 Q 1/68		G 0 1 N 33/53	M 4 B 0 6 3
G 0 1 N 33/53		37/00	1 0 2
37/00	1 0 2	C 1 2 N 15/00	Z N A F
審査請求 未請求 請求項の数7 O L (全 11 頁)			

(21) 出願番号 特願2001-142170(P2001-142170)

(22) 出願日 平成13年5月11日(2001.5.11)

(71) 出願人 000233055

日立ソフトウェアエンジニアリング株式会
社

神奈川県横浜市中区尾上町6丁目81番地

(72) 発明者 上野 紳吾

神奈川県横浜市中区尾上町6丁目81番地

日立ソフトウェアエンジニアリング株式会
社内

(74) 代理人 100091096

弁理士 平木 祐輔 (外1名)

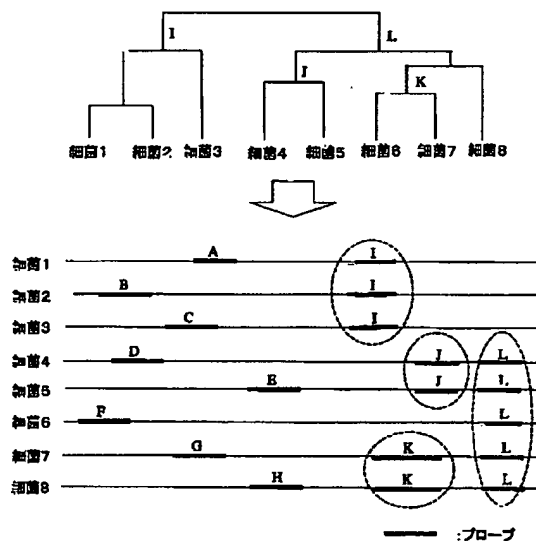
最終頁に続く

(54) 【発明の名称】 プローブ設計方法及びバイオチップ

(57) 【要約】

【課題】 バイオチップで生物種の識別を厳密に行う方法、種以上のレベルでの生物種の識別を可能にする方法を提供する。また、多数の生物種間での特異的プローブの選択を容易にする。

【解決手段】 ターゲットの系統樹に基づき、あるノード以下のリーフで共通した塩基配列があれば、それをそのノードに固有のプローブとして設計する。このプローブとリーフに固有のプローブをセットにして使用することにより、生物種の識別を厳密に行うことができ、また、種以上のレベルでの生物種の識別も可能になる。リーフに相当する塩基配列が同じもしくは似ていても、ノードに相当する配列が異なればプローブとして使用可能であるため、多数の生物種間での特異的プローブの選択が容易になる。



【特許請求の範囲】

【請求項1】 サンプル中に含まれる複数種類のターゲット生体高分子を判別するためのプローブの設計方法において、

前記複数種類のターゲット生体高分子を含む生体高分子群の分子系統樹のノードに対応するプローブとして、そのノード以下の生体高分子とのみ共通してハイブリダイズするプローブを設計することを特徴とするプローブの設計方法。

【請求項2】 サンプル中に含まれる複数種類のターゲット生体高分子を判別するためのプローブの設計方法において、

前記複数種類のターゲット生体高分子とそれぞれ特異的にハイブリダイズするプローブと、

前記複数種類のターゲット生体高分子を含む生体高分子群の分子系統樹のノードに対応するプローブとして、そのノード以下の生体高分子とのみ共通してハイブリダイズするプローブとを設計することを特徴とするプローブの設計方法。

【請求項3】 サンプル中に含まれる複数種類のターゲット生体高分子を判別するためのプローブの設計方法において、

前記複数種類のターゲット生体高分子を含む生体高分子群の分子系統樹のノードに対応するプローブとして、そのノード以下の生体高分子とのみ共通してハイブリダイズするプローブと、

前記ノード以下のターゲット生体高分子とそれぞれ特異的にハイブリダイズするプローブとを設計することを特徴とするプローブの設計方法。

【請求項4】 複数種類のターゲット生体高分子を判別するために基板上に複数のプローブをスポットしたバイオチップにおいて、

前記複数種類のターゲット生体高分子を含む生体高分子群の分子系統樹のノードに対応するプローブとして当該ノード以下の生体高分子とのみ共通してハイブリダイズするプローブをスポットしたことを特徴とするバイオチップ。

【請求項5】 複数種類のターゲット生体高分子を判別するために基板上に複数のプローブをスポットしたバイオチップにおいて、

前記複数種類のターゲット生体高分子とそれぞれ特異的にハイブリダイズするプローブと、前記複数種類のターゲット生体高分子を含む生体高分子群の分子系統樹のノードに対応するプローブとして当該ノード以下の生体高分子とのみ共通してハイブリダイズするプローブとをスポットしたことを特徴とするバイオチップ。

【請求項6】 複数種類のターゲット生体高分子を判別するために基板上に複数のプローブをスポットしたバイオチップにおいて、

前記複数種類のターゲット生体高分子を含む生体高分子

群の分子系統樹のノードに対応するプローブとして当該ノード以下の生体高分子とのみ共通してハイブリダイズするプローブと、前記ノード以下のターゲット生体高分子とそれぞれ特異的にハイブリダイズするプローブとをスポットしたことを特徴とするバイオチップ。

【請求項7】 プローブとのハイブリダイゼーション反応に基づいてターゲット生体高分子の存在を検出するターゲット検出方法において、

ターゲットとなるべき複数種類の生体高分子を含む生体高分子群に対する分子系統樹の所定のノード以下の生体高分子とのみ共通してハイブリダイズするプローブとのハイブリダイゼーション反応と、前記所定のノード以下の生体高分子とそれぞれ特異的にハイブリダイズするプローブとのハイブリダイゼーション反応とをもとにターゲット生体高分子の存在を検出することを特徴とするターゲット検出方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、サンプル中に含まれる複数種類のDNAを判別することを目的とするバイオチップ及びそのバイオチップにスポットするプローブの設計方法に関する。

【0002】

【従来の技術】近年の遺伝子解析技術の発展によって、遺伝子の機能や構造が次第に明らかになってきた。中でも、DNAチップまたはDNAマイクロアレイ（以下、本明細書ではバイオチップという）の技術は、遺伝子解析に有効な手段であるとして注目されている。バイオチップとは、ガラス、シリコン、プラスチックなどの基板上に多数の異なるプローブを高密度に整列化してスポットしたものである。プローブとしては、通常cDNAや、20～30mer程度の短鎖ヌクレオチドなどが用いられる。バイオチップの原理は、DNAを構成する4つの塩基A（アデニン）、T（チミン）、G（グアニン）、C（シトシン）において、AとT、GとCが互いに水素結合する性質、すなわちハイブリダイゼーションに基づいている。このバイオチップ上に蛍光物質などで標識したDNAやRNAなどのターゲットを注ぎ、プローブとハイブリダイゼーションさせることにより、ターゲットを捕獲する。捕獲されたDNAやRNAなどのターゲットは各プローブスポットからの蛍光シグナルとして検出され、これをコンピュータでデータ解析することにより、ターゲット中の数千から数万のDNAやRNAの状況を一挙に観測することが可能となる。

【0003】バイオチップの利用方法のひとつとして、ターゲットとなる遺伝子（またはDNA断片）を捕獲することで、調べたいターゲットのDNAがサンプル中に含まれているか検定したり、捕獲したDNAの配列を読み取ったり、SNPなどのDNAの多型部分を調べる利用法（SBH法）がある。

【0004】バイオチップの具体的な応用例の一つとし

て、生物種の識別方法について説明する。生物種の同定には通常サンプル中のターゲットDNAの16SrDNA領域を使うことが多い。16SrDNAはその生物に特異的なDNA配列で、かつ変異が起こりにくい領域であるので生物種の同定に使用するのに適しているからである。そして、プローブはその領域に合わせて設計する。実際の実験では16SrDNAを特異的に増幅させ、プローブとハイブリダイズさせることで生物種を識別していく。

【0005】図1は、バイオチップを用いた細菌同定の方法を模式的に示す説明図である。まず、細菌のDNA配列を格納したデータベース10から各細菌P、Q、R、…に特異的な16SrDNA領域の塩基配列をプローブ11、12、13、…として選択し、プローブ設計を行う。そのプローブ設計に従って各細菌に対応するプローブを作製し、それを基板上に縦横に並べてスポットし、バイオチップ14を作製する。そのバイオチップ14上に、患者の血液、痰などから抽出し蛍光標識したDNAをターゲット15として注ぎ、バイオチップ14上のプローブとハイブリダイゼーションさせる。その結果、図の中央に示すように、(横No1:縦No2)のスポットと(横No3:縦No5)のスポットからシグナルが観測されたとする。このとき、バイオチップ上のスポット位置と細菌との対応表から細菌(*Actinobacillus actinomycetemcomitans*)と(*Klebsiella oxytoca*)がターゲットに混入している(可能性がある)ことがわかる。この場合、検出されるシグナルと菌種とは一対一の関係にある。

【0006】

【発明が解決しようとする課題】従来のバイオチップのプローブ設計法は、ターゲットとプローブとを一対一に対応づけるものである。しかし、このようなプローブ設計法は必ずしも満足のいくものではなかった。まず第1に、生物種のDNAとプローブが一対一の関係では、突然変異や実験誤差によって生物種の判定が厳密に行えない場合がある。

【0007】実験誤差の例として、ターゲットのDNA断片が、バイオチップ上の対応する相補的な配列のプローブと結合しなかったり、あるいは対応しないプローブと結合したりすることがある。対応するプローブと結合しない場合の例として、ターゲットのDNA配列がプローブ設計時に参照したパブリックデータベースの配列とは異なることがある。例えば図2に示すように、バイオチップに注ぐターゲット22のDNAに変異が起きていて、1塩基置換や1塩基挿入が起きていると、プローブ21とハイブリダイズしない。図2は、ターゲット22に位置23で1塩基挿入があって結合しない場合を示している。

【0008】対応しないプローブと結合する場合のひとつとしては、クロスハイブリダイゼーションがある。クロスハイブリダイゼーションとは、図3に示すように、ターゲットの遺伝子(またはDNA断片)33、34と、

バイオチップ上のプローブ31、32とが類似したDNA配列を持つとき、部分的に結合される状態をいう。文献(Michael D. Kane et al.: Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays: *Nucleic Acids Res.*, 28(22), 4552-4557, 2000)によると、配列の類似度が75%以上のもの、あるいは類似度がそれほど高くない場合(50~75%)でも、長さ15mer以上の連続した相補文字列があれば、クロスハイブリダイゼーションする可能性があるとの報告がある。

【0009】一方、クロスハイブリダイゼーションを起こさないようにするための試みとして、配列に特異的なプローブを選別する方法(Ken-ichi Kurata et al.: *ProbeDesign for DNA Chips: Genome Informatics* 1999, 225-6, 1999)などがあるが、確実にクロスハイブリダイゼーションしないレベルには至っていない。また、ある程度のクロスハイブリダイゼーションが起きているとして、そこから本来の蛍光シグナルがどの程度あるかを予測する方法(Mitsuteru Nakao et al.: *Quantitative Estimation of Cross-Hybridization in DNA Microarrays Based on a Linear Model: Genome Informatics* 2000, 231-232, 2000)も考えられているが、これも実用レベルには至っていない。

【0010】クロスハイブリダイゼーション以外にも、ハイブリダイゼーション反応のときの温度やターゲット溶液のpHといった実験条件、実験機器の状態、ターゲットやプローブの濃度などにより、本来該当しないスポットから蛍光シグナルが観測されてしまう可能性はいくつもある。現在のバイオチップでは、同じDNA配列のプローブをチップ上に複数個スポットすることによって、実験に再現性があるかを確かめてはいるが、この方法では上に述べた実験誤差に対応することが出来ない。

【0011】第2に、ターゲットとプローブとを一対一に対応づけたバイオチップでは、種以上のレベルでの生物種の同定ができない。従来の生物種同定チップでは、生物の種ではなく属や科のレベルで分類を行いたい場合のような大まかなレベルの検出に対する要望には応えることができなかった。例えば、ある生物の特徴が種レベルではほとんど差がないため属レベルで生物の分類を行いたい場合、従来のバイオチップでは対応できない。

【0012】第3に、多数の生物種それぞれに対して特異的なプローブを選出しようすると、生物種が多くなるにつれ、特異的なプローブの選出に限界が生じてくる。図4は、例えば50個のプローブを選出しようとしたとき、選出されたプローブNo. 1~No. 50の中にDNA配列が互いに似ているプローブが多くなり、プローブの選出が非常に困難になる状況を模式的に示している。また、配列が類似していること以外にも、たくさんプローブをとるとプローブ間の T_m 値が揃わないといった問題もある。 T_m 値は2本鎖のDNAが解離して1本鎖になる温度

である。ハイブリダイゼーション反応は、高温下では2本鎖DNAが解離して1本鎖になり、低温下では再び2本鎖を形成するDNAの性質を利用している。従って、バイオチップにおいてはスポットするプローブのTm値を均一にする必要がある。

【0013】本発明は、このような従来技術の問題点に鑑み、ターゲットの遺伝子（またはDNA断片）をより高精度かつ確実に検出することのできるバイオチップ及びプローブ設計方法を提供することを目的とする。また、本発明は、大まかなレベルでの生物種の同定を可能にするバイオチップ及びそのプローブ設計方法を提供することを目的とする。さらに、本発明は、多数の生物種間での種特異的プローブの選出を容易にするプローブ設計方法を提供することを目的とする。

【0014】

【課題を解決するための手段】上記目的を達成するために、本発明では、分子系統樹による生物の分類に基づいた複数の特徴的なプローブを設計することで、ターゲット中のDNAがどの生物種由来のものであるかを判定することと、多数の生物種間での種特異的プローブの選出を容易にする。ここで、分子系統樹とは、生物種の生体高分子配列の相同性に基づいて作られた樹状図のことであり、同一のノード以下に分類される生物種は近縁であり生物学的に似た性質を持つ。

【0015】複数の特徴的なプローブを設計する際の方針は、従来のような生物種との一対一の関係だけでプローブを設計するのではなく、いくつかのターゲットに共通するDNA配列をプローブとして選択することである。その際、分子系統樹を入力データとして各ノードに対応するプローブを設計していく。つまり、分子系統樹のあるノード以下の細菌に共通し、かつ、他の細菌にはない塩基配列があれば、それをそのノードに固有のプローブとして設計する。

【0016】図5は、分子系統樹に基づいてプローブを設計した例を示す説明図である。図のI, J, K, Lに対応するノードに固有のプローブがとれた状況を示している。図5の例の場合、塩基配列Iは、分子系統樹のノードI以下の細菌1, 2, 3に共通し、しかも他の細菌にはない塩基配列である。従って、ノードIに対応するプローブとして塩基配列Iを選定する。同様に、塩基配列Jは、分子系統樹のノードJ以下の細菌4, 5に共通し、他の細菌にはない塩基配列であるため、ノードJに対応するプローブとする。塩基配列Kは、分子系統樹のノードK以下の細菌7, 8に共通し、他の細菌にはない塩基配列であるため、ノードKに対応するプローブとする。塩基配列Lは、分子系統樹のノードL以下の細菌4～8に共通し、他の細菌にはない塩基配列であるため、ノードLに対応するプローブとする。なお、塩基配列A, B, C, ..., Hは、細菌1～8に固有の塩基配列であり、細菌1～8に一対一対応する各細菌に固有のプロ

ーブとして選出される。

【0017】このように分子系統樹のノードに対応するプローブを設計すると、検出されたスポットの菌名だけでなく、それらの近縁関係がわかるため、ターゲットに含まれている細菌をより精密に識別することができる。実際、分子系統樹はDNA配列の相同性に基づいて作られているが、形態学的に作製された進化系統樹とほぼ一致する。そのため、進化系統樹に基づく種や属といった分類法は分子系統樹のノードとリーフの関係と一致することが多い。

【0018】また、図5のように分子系統樹による分類に準じたプローブを用意するプローブ設計法は、1つのターゲットに対して固有なプローブを数種類用意する方法に対して、バイオチップ上に設けるスポットの数量も削減でき、コスト面と労力面共にメリットがある。そして、単に細菌に特異的なプローブを複数用意することによって、多くのスポットからのシグナルで細菌の混入具合を判別するので、より正確な判定を行うことが出来る。

【0019】図6は、図5で選択したプローブA～Kをスポットしたバイオチップにターゲットをハイブリダイゼーションさせたときの結果の例を示す模式図である。図6に示した例では、プローブA, C, G, I, K, Lから信号が検出されており、細菌1、細菌3、細菌7が混入していることが、細菌に固有なプローブ（A, C, G）と、分子系統樹の中間ノードに対応するプローブ（I, K, L）から判断することが出来る。このように中間ノードからのシグナルを利用することにより精度の高い検出が可能になる。

【0020】また、図7のように、種に相当するスポットGからのシグナルが検出されているにもかかわらず中間ノードに相当するスポットK, Lからのシグナルが検出されていない場合は、種に相当するスポットGでクロスハイブリが起きていると考えられる。つまり、中間ノードに相当するスポットによりハイブリ反応が正常に行われたかどうかを判別できる。このような検出手法を用いることにより一層精度の高い検出が可能となる。また逆に、種に相当するスポットA, B, Cからのシグナルが検出されず、中間ノードに相当するスポットIからのシグナルが検出されている場合は、未知種または突然変異した種のDNAがサンプル中に存在していると考えられる。この場合は、種レベルまでは識別できなくてもその上位レベルまでは識別することができ、未知のものを推定する手がかりとなる。

【0021】菌種を同定するプローブを各菌の16SrDNA配列から選定するとき、それぞれのプローブは互いに似ていてはいけない。その結果、菌種が増加してくると互いに似ていない塩基配列を選ぶことは困難になってくる。しかし、図8のように、種に相当する塩基配列が同じもしくは似ていても、中間ノードに相当する配列が異

なれば、中間ノードに相当する配列と組み合わせること
でプローブとして使用可能である。図8の例では、細菌
1～3は α 属に、細菌48～50は β 属に属している。
プローブNo.1とNo.49は非常に良く似た配列である。
このような状況でも、属に相当するプローブ α や β を種
に相当するプローブと同時に使用することにより、本来
なら互いに非常に似ているため使用できない配列No.1
とNo.49も種のプローブとして利用可能になる。ター
ゲットの検出に当たっては、図6にて説明したように
種、中間ノードそれぞれに相当する複数のプローブから
のシグナルから総合的に判断する。

【0022】本発明によるプローブの設計方法、バイオ
チップ、ターゲット検出方法の特徴をまとめると、以下
の通りである。

(1) サンプル中に含まれる複数種類のターゲット生体
高分子を判別するためのプローブの設計方法において、
前記複数種類のターゲット生体高分子を含む生体高分子
群の分子系統樹のノードに対応するプローブとして、そ
のノード以下の生体高分子とのみ共通してハイブリダイ
ズするプローブを設計することを特徴とするプローブの
設計方法。

【0023】(2) サンプル中に含まれる複数種類のター
ゲット生体高分子を判別するためのプローブの設計方
法において、前記複数種類のターゲット生体高分子とそ
れぞれ特異的にハイブリダイズするプローブと、前記複
数種類のターゲット生体高分子を含む生体高分子群の分
子系統樹のノードに対応するプローブとして、そのノード
以下の生体高分子とのみ共通してハイブリダイズするプ
ローブとを設計することを特徴とするプローブの設計方
法。

【0024】(3) サンプル中に含まれる複数種類のター
ゲット生体高分子を判別するためのプローブの設計方
法において、前記複数種類のターゲット生体高分子を含
む生体高分子群の分子系統樹のノードに対応するプロー
ブとして、そのノード以下の生体高分子とのみ共通して
ハイブリダイズするプローブと、前記ノード以下のターゲ
ット生体高分子とそれぞれ特異的にハイブリダイズする
プローブとを設計することを特徴とするプローブの設計
方法。

【0025】(4) 複数種類のターゲット生体高分子を
判別するために基板上に複数のプローブをスポットした
バイオチップにおいて、前記複数種類のターゲット生体
高分子を含む生体高分子群の分子系統樹のノードに対応
するプローブとして当該ノード以下の生体高分子とのみ
共通してハイブリダイズするプローブをスポットしたこ
とを特徴とするバイオチップ。

【0026】(5) 複数種類のターゲット生体高分子を
判別するために基板上に複数のプローブをスポットした
バイオチップにおいて、前記複数種類のターゲット生体
高分子とそれぞれ特異的にハイブリダイズするプローブ

と、前記複数種類のターゲット生体高分子を含む生体高
分子群の分子系統樹のノードに対応するプローブとして
当該ノード以下の生体高分子とのみ共通してハイブリダ
イズするプローブとをスポットしたことを特徴とするバイ
オチップ。

【0027】(6) 複数種類のターゲット生体高分子を
判別するために基板上に複数のプローブをスポットした
バイオチップにおいて、前記複数種類のターゲット生体
高分子を含む生体高分子群の分子系統樹のノードに対応
するプローブとして当該ノード以下の生体高分子とのみ
共通してハイブリダイズするプローブと、前記ノード以下
のターゲット生体高分子とそれぞれ特異的にハイブリダ
イズするプローブとをスポットしたことを特徴とするバイ
オチップ。

【0028】(7) プローブとのハイブリダイゼーショ
ン反応に基づいてターゲット生体高分子の存在を検出す
るターゲット検出方法において、ターゲットとなるべき
複数種類の生体高分子を含む生体高分子群に対する分子
系統樹の所定のノード以下の生体高分子とのみ共通して
ハイブリダイズするプローブとのハイブリダイゼーション
反応と、前記所定のノード以下の生体高分子とそれぞれ
特異的にハイブリダイズするプローブとのハイブリダイ
ゼーション反応とをもとにターゲット生体高分子の存在
を検出することを特徴とするターゲット検出方法。

【0029】

【発明の実施の形態】以下、本発明を実施する場合の一
形態を図面を参照して具体的に説明する。図9は、バイ
オチップの作製、蛍光シグナルの検出、シグナルデータ
の解析を行うバイオチップシステムの構成例を示すブロ
ック図である。このバイオチップシステムは、配列デー
タの入出力及び実験データの解析等を行う中央処理装置
900、中央処理装置900での処理に必要なプログラムを
格納するプログラムメモリ910、キャラクタ及び
グラフィック画面を表示する表示装置901、本システ
ムへの値の入力や選択の操作を行うためのキーボード9
02およびマウス903、プローブDNA配列の設計に使
うターゲットDNAの情報が格納されている配列データベ
ース904、ノード用プローブの設計に使う系統樹の情
報が格納されている系統樹データ909を備える。

【0030】ここで、配列データベース904はローカ
ルなデータベースであってもよいし、ネットワーク等を
介して遠隔地に設置されたサーバーコンピュータが管理
しているデータベースであってもよい。系統樹データ9
09は既に作成されたデータであってもよいし、新たに
配列データベース904から作成したデータであっても
よい。また、系統樹データ909はローカルなコンピュ
ータにあるデータであってもよいし、ネットワーク等を
介して遠隔地に設置されたサーバコンピュータが管理し
ているデータであってもよい。中央処理装置900はコ
ンピュータとそのプログラムによって具体化されるもの

である。

【0031】プログラムメモリ910は、配列データベース904のデータを処理する配列データ処理部911、系統樹データ909を解析する系統樹データ解析処理部912、キーボード902やマウス903からの入力を処理する入力データ処理部911、配列データ処理部911の処理結果と系統樹データ解析処理部912の解析結果をもとにプローブの選択処理を行うプローブ選択処理部914、設計したプローブを表示するプローブ表示処理部915からなる。

【0032】中央処理装置900は、設計したプローブDNA配列から実際のチップに載せるプローブを作製するプローブ作製装置905、プローブ作製装置で作製したプローブを入れるウェル906からプローブを取り出してチップ908上の所定の位置に載せるスポット907の制御も行う。

【0033】図10は、本システムが管理するターゲットDNA配列データの例を示したものである。配列情報は、dnaSeq[i] (i = 1, 2, ..., sNum) という長さsNum個の構造体の配列に格納する。ただしsNumはプローブ選別で対象とするターゲットDNAの個数である。配列dnaSeq[] は、配列名(1000)、DNA配列(1001)、DNA配列の配列長(1002)、そしてこの配列がどのプローブで検出されるのかを示すPROBE_ID(1003)からなる。PROBE_IDには、このターゲットDNAを識別することが可能な各プローブの識別子が入る。この識別子は後述する配列probe[]の識別子を指している。また、これらの属性に加えて、配列を取り出した生体組織(器官)の名前、生物種名、属名、配列データベースに関する情報などをdnaSeq[]の属性として加えてもよい。

【0034】図11は、本システムが管理するプローブDNA配列データの例を示したものである。プローブDNA配列データは、probe[i] (i = 1, 2, ..., pNum) という長さpNum個の構造体の配列に格納されている。ここでpNumは、チップ上に載せるプローブの総数である。配列probe[] は、チップ上でのプローブの座標位置(1100)、検出器で観測された蛍光シグナル強度(1101)、プローブのDNA配列(1102)、このプローブによって検出することが可能なターゲットのリストを表すTARGET_ID(1103)からなる。TARGET_IDには、前に述べた配列dnaSeq[]のインデックスをターゲットの識別子として入力する。

【0035】図12は、本システムの入力データである系統樹データの例を示したものである。系統樹データはファイル形式になっており、系統樹のリーフの部分をdnaSeq[]の識別子に対応させ、一組の括弧を一つの中間ノードに対応させている。そして中間ノードが(系統樹上でリーフに近い)中間ノードを持つときは、入れ子構造で表現する形式をとっている。すなわち、BNF記法で書く次のようになる。

【0036】

ノード ::= (ノード, ノード) | dnaSeq[] の識別子
そして、系統樹データには、このルートに対応するノードがかかっている。図12で示した系統樹データの例では、(1, 2)はノードAに対応しており、((1, 2), 3)はノードBに対応している。

【0037】図13は、本システムが管理するNode構造体を示す図である。Node構造体は系統樹における各ノードおよびリーフを示したものである。Nodeはリーフ識別子1300、左子ノードへのポインタ1301、右子ノードへのポインタ1302からなる。リーフの識別子1300にはNodeが系統樹における中間ノードのときNodeの下に属するリーフの識別子(dnaSeq[])のインデックスが登録され、Node自体がリーフのときは、リーフの識別子1300には対応するdnaSeq[]のインデックスが登録される。また、Nodeがリーフのときは、左子ノードのポインタ、右ノードのポインタにはNULLを入れる。

【0038】図14はNode構造体の間の関係を示したもので、左子ノードと右子ノードへのポインタをつなぎ合わせることで系統樹のツリー構造を再生している。図15は、本発明の概略処理フローを示した図である。まず、配列データベース904からプローブ選別で対象とするターゲットDNA配列データを読み込み、dnaSeq[]に登録する(ステップ1500)。次に、系統樹データ909から系統樹データを読み込み、Node構造体に登録する。系統樹データ909は既に作成されたデータであってもよいし、新たに配列データベースから作成したデータであってもよい。入力された系統樹データは、図14のように樹状図の形状に合わせて、Node構造体のリンクを構築していく(ステップ1501)。

【0039】次に、プローブ選別の基準を、キーボード902やマウス903を用いて入力する。すなわち、何merのプローブを作製するのか、プローブのTm値(DNAの2本鎖が1本鎖に分離する温度)、他のターゲットDNAとの配列類似度の限界値など、プローブDNA配列選別の要件に関する情報を設定していく。入力方法は他にも、予めプローブ作製に関する情報が入っているファイルを読み込む利用形態がある(ステップ1502)。その後、dnaSeq[]及びNodeを用いて、系統樹のノード及び種に対応するプローブDNA配列を決める(ステップ1503)。この処理については後で詳しく述べる。この処理によりプローブが配列probe[i] (i = 1, ..., pNum)に格納される。

【0040】これをプローブ作製装置905に送り、実際にプローブを作る(ステップ1504)。作ったプローブをウェル906に調整し、ウェルからスポット907でバイオチップを作製する(ステップ1505)。最後に樹状図に対応したプローブの選別結果を図17のように表示装置901に表示する(ステップ1506)。図17については後で詳しく述べる。

【0041】図16は、図15におけるプローブDNA配列を決める処理（ステップ1503）の詳細フローである。図15のステップ1503では、引数として系統樹のルートを与えてコールする。図16において、まず、引数として与えられたNode構造体のデータを読み込む（ステップ1600）。次に、このNodeに子ノードが存在するかどうか調べる（ステップ1601）。子ノードが存在しないのであれば、Nodeは系統樹の種に対応している。存在するのであれば、系統樹のノードに対応している。

【0042】Nodeに子ノードが存在しないとき、まず、このNodeのリーフの識別子メンバ1300に対応するターゲットに対するプローブDNA配列を選別する。ターゲットDNA配列から、プローブの長さ分のDNA部分配列（プローブ候補）を取り出し、プローブ候補が、全DNA配列でユニークであるか、基準となるTm値を満たしているか、他のDNA配列との配列類似度が限界値を超えていないか、クロスハイブリを起しやすい配列ではないか、などを調べる。これらの基準を満たし、かつ最も望ましいプローブ候補をターゲットDNAに固有なプローブとして選別する。そして、選別したプローブDNA配列をprobe[]のDNA配列1102に登録し、Nodeのリーフの識別子メンバをTARGET_ID1103に追加する（ステップ1602）。Nodeのリーフ識別子メンバに対応するdnaSeq[]のPROBE_IDメンバ1003に選別したプローブの識別子を追加する（ステップ1603）。

【0043】ステップ1601において、Nodeに子ノードが存在するとき、まず、このNodeに対応するプローブDNA配列を選別する。Nodeに対応するプローブは、Nodeの下の子種とは反応し、それ以外の種とは反応しないものでなくてはならない。そこで、まず、Nodeのリーフの識別子メンバで示される識別子のターゲットDNA配列には含まれ、それ以外のターゲットDNA配列には含まれないような、プローブの長さ分の部分配列をプローブ候補として求める。その後、Tm値や配列類似度は基準を満たしているかなどを調べ、最も望ましいプローブ候補をDNA配列に固有なプローブとして選別する。選別したプローブDNA配列をprobe[]のDNA配列1102に登録し、Nodeのリーフの識別子メンバをTARGET_ID1103に追加する（ステップ1604）。Nodeのリーフ識別子メンバに対応するdnaSeq[]のPROBE_IDメンバ1003に選別したプローブの識別子を追加する（ステップ1605）。

【0044】その後、引数をNodeの左右の子ノードとして、ステップ1600から処理を繰返す（ステップ1606、1607）。このようにして、系統樹の全てのノード及び種を巡回しながらプローブを選別していく。また、望ましいプローブ候補が得られなかったときは、そのことを表示装置901に出力する。

【0045】図17は、本システムが選別したプローブの情報を表示する表示装置901の画面の例を示す図で

ある。系統樹データ909が読み込まれ表示画面1700に表示されると、マウス903のカーソル1701を用いて系統樹のノードを選択する。ノードの選択は、マウス903の他、キーボード902を用いて行っても良い。すると1702、1703、1704、1705が表示される。1702は、マウスカーソル1701で選択したノードに属する生物種（ここではStr. sanguiniとStr. CanisとEnt. aviumの3つ）のマルチプルアライメントの結果である。網掛けの部分は、三つの生物種でDNA配列が一致する部分である。網掛けでない部分は、少なくとも一つの生物種でDNA配列が異なる部分である。1703は、カーソル1701で選択したノードに対応するプローブの一つである。1704は、そのプローブのDNA配列中における位置を示す。配列1703は7塩基目から始まるので、マルチプルアライメントの7塩基目から表示される。1705は、カーソル1701で選択したノードに対応するプローブの一覧表である。ここではプローブの番号、配列、DNA配列中における位置、反応温度を表示しているが、自分自身で絡み合う度合いや他の条件を表示しても良い。以上の処理によって、調べたいターゲットDNAの由来する生物種の判別を対象としたプローブをうまく選別することが可能となる。

【0046】

【発明の効果】本発明によると、ターゲットの遺伝子（又はDNA断片）を高精度に、あるいは所望の分類レベルで検出することのできるバイオチップを得ることができる。また、調べたいターゲットDNAの種類が増加してもプローブの選別が容易に行える。それぞれのプローブは系統樹のノードとリーフの関係に対応しているので、ハイブリダイゼーション反応やシグナル読み取りにおけるエラーチェックの役割も果たす。

【図面の簡単な説明】

【図1】バイオチップを用いた細菌同定の方法を模式的に示す説明図。

【図2】ターゲットDNAが対応するプローブと結合しない例を示す図。

【図3】ターゲットDNAが対応しないプローブと結合する例を示す図。

【図4】菌種が多数の場合、プローブの選別が困難になることを説明する図。

【図5】分子系統樹に対応したプローブの設計例を示す図。

【図6】本発明のバイオチップを用いた実験結果の例を示す図。

【図7】本発明のバイオチップを用いた実験結果の他の例を示す図。

【図8】種に相当する塩基配列が同じもしくは似ていても、中間ノードに相当する配列が異なればプローブとして使用可能であることを示す図。

【図9】本発明によるバイオチップシステムの構成例を示すブロック図。

【図10】ターゲットDNA配列データのデータ構造を示す図。

【図11】プローブDNA配列データのデータ構造を示す図。

【図12】系統樹データの構造を示した図。

【図13】Node構造体のデータ構造を示した図。

【図14】Node構造体のリンク関係を示した図。

【図15】本発明の概略処理フローを示す図。

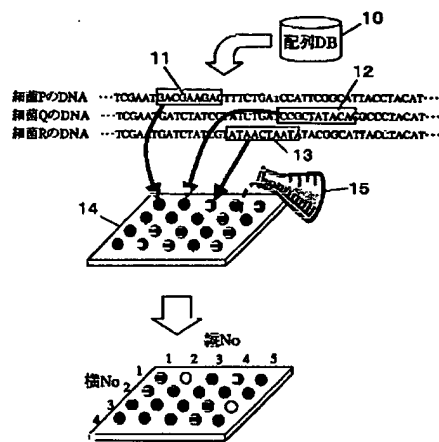
【図16】プローブ配列決定の詳細フローを示す図。

【図17】プローブ選別結果表示画面例を示す図。

【符号の説明】

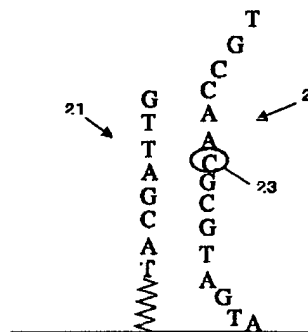
10…配列データベース、11～13…プローブ、14…バイオチップ、15…ターゲット、21…プローブ、22…ターゲット、31、32…プローブ、33、34…ターゲット、900…中央処理装置、901…表示装置、902…キーボード、903…マウス、904…配列データベース、909…系統樹データ、910…プログラムメモリ

【図1】

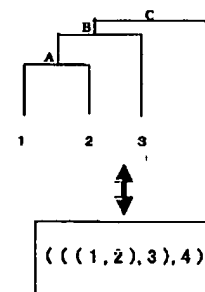


検体No.	検体名	菌名
1	1	Acinetobacter calcoaceticus (細菌P)
1	2	Actinobacillus actinomycetemcomitans (細菌Q)
1	3	Aeromonas hydrophila (細菌R)
1	4	Campylobacter fetus
1	5	Clostridium funduli
2	1	Clostridium perfringens
2	2	Enterobacter aerogenes
2	3	Enterococcus avium
2	4	Enterococcus casseliflavus
2	5	Enterococcus faecalis
3	1	Enterococcus gallinarum
3	2	Escherichia coli
3	3	Flavobacterium indologenes
3	4	Haemophilus influenzae
3	5	Klebsiella oxytoca
4	1	Klebsiella pneumoniae
4	2	Listeria monocytogenes
4	3	Mycobacterium abscessus
4	4	Mycobacterium avium
4	5	Mycobacterium goodii

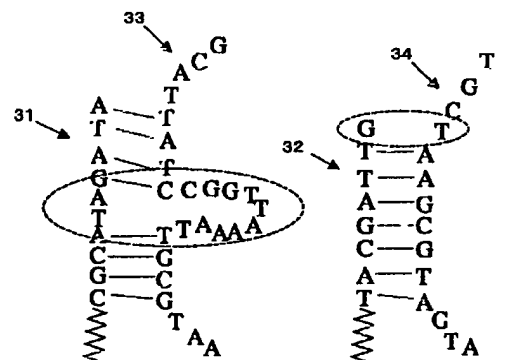
【図2】



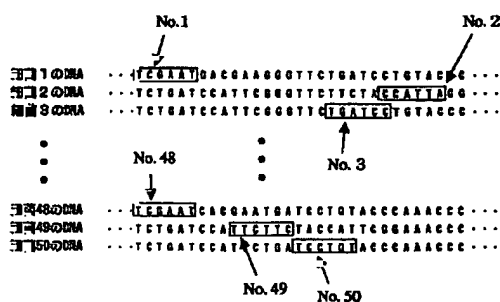
【図12】



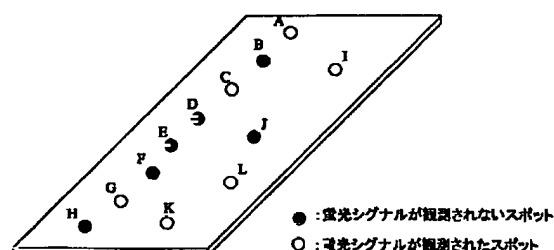
【図3】



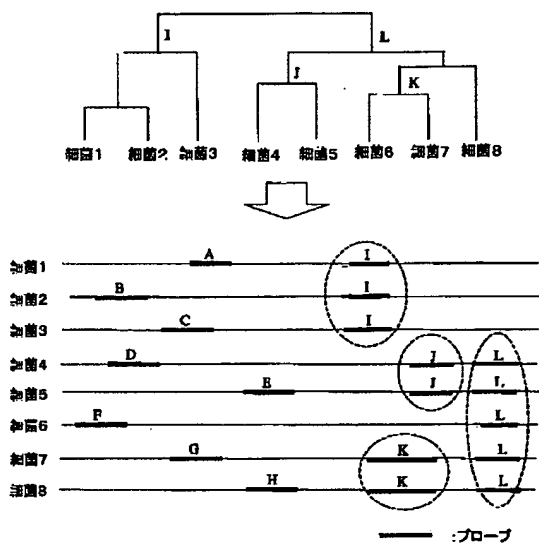
【図4】



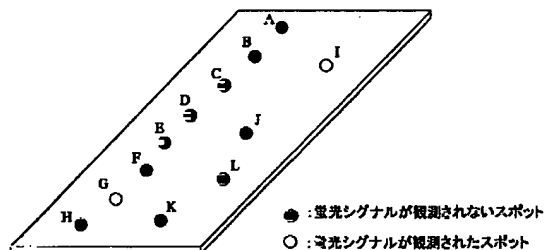
【図6】



【図5】



【図7】

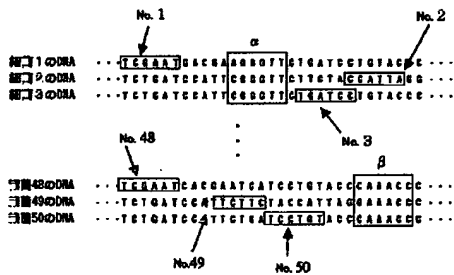


【図10】

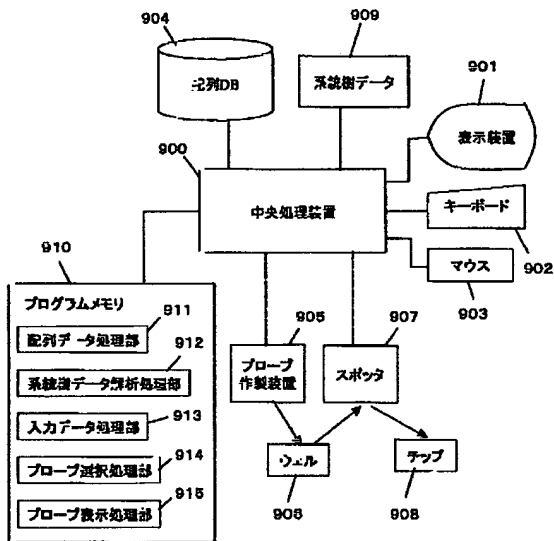
dnaSeq[i] (i = 1, 2, ..., Num)

メンバー名	値
1000 配列名	Escherichia coli
1001 DNA配列	aaatgaga gttgatct ggtcagatt.....
1002 配列長	1,541
1003 PROBE_ID	1,2,3

【図8】



【図9】



【図11】

probe[i] (i = 1, 2, ..., Num)

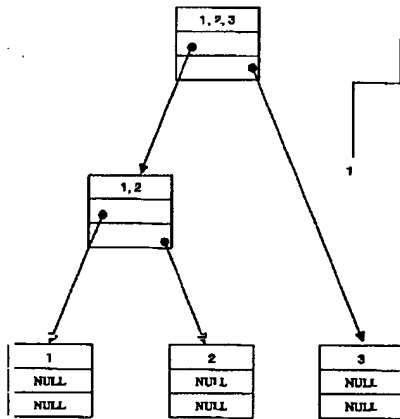
メンバー名	値
1100 座標位置	2,3
1101 蛍光シグナル強度	1,923
1102 DNA配列	aaatgaga gttgatct
1103 TARGET_ID	1,2,3

【図13】

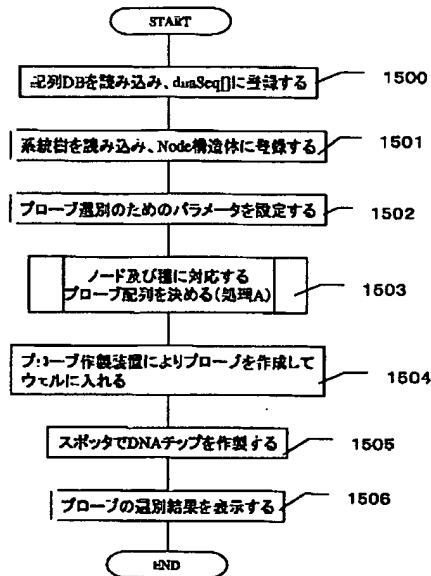
Node

メンバー名	値
1300 リーフの識別子	1,2,3
1301 左子ノードへのポインタ	—
1302 右子ノードへのポインタ	—

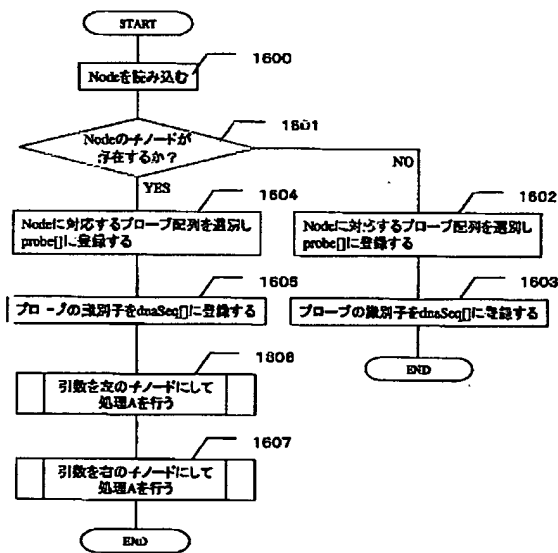
【図14】



【図15】

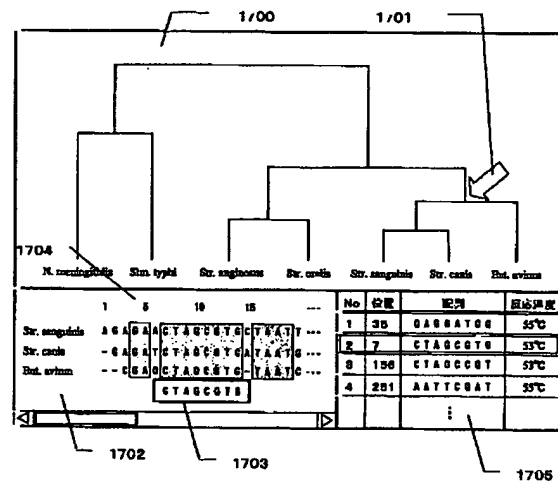


【図16】



【図17】

プローブ選別結果表示画面



フロントページの続き

(72)発明者 中重 亮
神奈川県横浜市中区尾上町6丁目81番地
日立ソフトウェアエンジニアリング株式会社
社内

(72)発明者 野崎 康行
神奈川県横浜市中区尾上町6丁目81番地
日立ソフトウェアエンジニアリング株式会社
社内

(特 1) 102-330768 (P2002-330768A)

(72)発明者 松本 俊子
神奈川県横浜市中区尾上町6丁目81番地
日立ソフトウェアエンジニアリング株式会
社内

(72)発明者 田村 卓郎
神奈川県横浜市中区尾上町6丁目81番地
日立ソフトウェアエンジニアリング株式会
社内

Fターム(参考) 4B024 AA19 AA20 CA01 CA09 CA11
HA11 HA12
4B029 AA07 AA23 BB20 FA15
4B063 QA13 QQ05 QQ42 QR08 QR42
QR56 QS25 QS34 QS39 QX02